

Subjective Logic with Uncertain Partial Observations

Lance M. Kaplan
US Army Research Laboratory
Adelphi, MD 20783-1197
lkaplan@ieee.org

Supriyo Chakraborty
UCLA
Los Angeles, CA 90095-1594
supriyo@ee.ucla.edu

Chatschik Bisdikian
IBM Research
Hawthorne, NY 10532
bisdik@us.ibm.com

Abstract—Subjective logic (SL) is an effective tool to manage and update beliefs over a set of mutually exclusive assertions. The method to update subjective beliefs from direct observations of assertions is well understood. Recent work has incorporated the SL framework to derive the belief update equations for partial observations where the measurements are only statistically related to the assertions. This work further expands the notion of SL to consider uncertainty in the underlying statistical relationship between measurements and assertions. In other words, new methods are derived for SL that incorporate uncertainty in the reported likelihood of the assertions. Simulations demonstrate the utility of the new likelihood uncertainty aware belief update methods.

I. INTRODUCTION

This work investigates methods to compensate for uncertain observations in updating subjective logic (SL) beliefs. SL has emerged as a rigorous method to represent and reason over human generated or automated beliefs in face of uncertainty [1]. Applications of SL include trust management [2], Bayesian networks [3], and fusion [1], [4]. In short, SL provides effective tools to manage and combine beliefs over a set of mutually exclusive assertions from multiple human or computer agents. At a given point in time, an agent's belief is the result of a prior belief and a set of observations. The uncertainty of the belief represents the reliance on the prior, and the uncertainty decreases as the agent incorporates more observations to form the beliefs over the set of assertions.

To our knowledge, the current SL operations focus on fusing beliefs or exploiting belief for inference. This work is concerned about how to update subjective beliefs from observations. Implicitly, SL provides the operations to update beliefs when one of the possible assertions is completely visible in the observation. In recent work, we expanded the notions of SL to incorporate partial observations where the assertions are only statistically related to the observations [5]. This work expands [5] by considering uncertainty in the knowledge about the the statistical relationship between the observations and the assertions.

Research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defense and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defense or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute for Government purposes notwithstanding any copyright notation hereon.

To make these notions a little more concrete, let us consider a motivating example where one wants to understand the criminal activity within a city. Specifically, one wants to understand if a crime happens, what is the probability that the crime occurs in any one of the districts. Without any initial data, one might look at socio-economic factors to develop an initial set of probabilities. Over time, one can log where a crime occurs and start to use these observations to update the probabilities. Clearly, as more observations are logged, the certainty associated with the generated probabilities increases. SL is well suited to infer the probabilities of a crime occurring in the districts and the uncertainty associated to these probabilities.

Now, let us assume that one is interested in where criminals live. The question is now when a crime occurs, what is the probability that the perpetrator lives in a particular district. Like before, one can start with a prior set of probabilities based upon the socio-economics factors. Furthermore, when a crime occurs, the location of the crime is readily available in the police report. However, the identity of the perpetrator may or may never be discovered. Therefore, it is generally not possible to log where the perpetrator lives. Sometimes, this information can be determined with great likelihood when the criminal is caught. Most likely, one only incorporates statistical models that link the probability of where the perpetrator lives conditioned on where a crime occurs. For instance, a criminal may not operate in his/her immediate neighborhood where he/she can easily be identified, and a criminal may not want to venture too far away either. This contextual information can help answer the questions of the distribution of criminals over the various districts within a city. This scenario is an example of a geospatial abduction problem (GAP) [6]. SL is suited to tackle such applications, but the notions of how to incorporate statistical (and not just hard) evidence of the appearance of an assertion (the home district of a perpetrator) need to be developed within the SL framework.

SL is a probabilistic logic for assigning and updating basic belief assignments (BBA). Classic SL considers BBAs on a set of mutually exclusive singletons [1], [2] to form an SL opinion. The attractive feature of SL is that the multinomial opinion has a one-to-one mapping with parameters of a Dirichlet distribution. Formally, the Dirichlet distribution is the conjugate prior of the multinomial distribution. This means that it is the natural distribution to represent knowledge about the weights

| Report Documentation Page | | | Form Approved OMB No. 0704-0188 | | |
|--|------------------------------------|--|---|---|---------------------------------|
| Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. | | | | | |
| 1. REPORT DATE JUL 2012 | | 2. REPORT TYPE | | 3. DATES COVERED 00-00-2012 to 00-00-2012 | |
| 4. TITLE AND SUBTITLE Subjective Logic with Uncertain Partial Observations | | 5a. CONTRACT NUMBER | | | |
| | | 5b. GRANT NUMBER | | | |
| | | 5c. PROGRAM ELEMENT NUMBER | | | |
| 6. AUTHOR(S) | | 5d. PROJECT NUMBER | | | |
| | | 5e. TASK NUMBER | | | |
| | | 5f. WORK UNIT NUMBER | | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory, Adelphi, MD, 20783-1197 | | 8. PERFORMING ORGANIZATION REPORT NUMBER | | | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | 10. SPONSOR/MONITOR'S ACRONYM(S) | | | |
| | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) | | | |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited | | | | | |
| 13. SUPPLEMENTARY NOTES Presented at the 15th International Conference on Information Fusion held in Singapore on 9-12 July 2012. Sponsored in part by Office of Naval Research and Office of Naval Research Global. | | | | | |
| 14. ABSTRACT Subjective logic (SL) is an effective tool to manage and update beliefs over a set of mutually exclusive assertions. The method to update subjective beliefs from direct observations of assertions is well understood. Recent work has incorporated the SL framework to derive the belief update equations for partial observations where the measurements are only statistically related to the assertions. This work further expands the notion of SL to consider uncertainty in the underlying statistical relationship between measurements and assertions. In other words, new methods are derived for SL that incorporate uncertainty in the reported likelihood of the assertions. Simulations demonstrate the utility of the new likelihood uncertainty aware belief update methods. | | | | | |
| 15. SUBJECT TERMS | | | | | |
| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT Same as Report (SAR) | 18. NUMBER OF PAGES 8 | 19a. NAME OF RESPONSIBLE PERSON |
| a. REPORT unclassified | b. ABSTRACT unclassified | c. THIS PAGE unclassified | | | |

associated to a loaded die after observing a number of dice rolls. The parameters of the Dirichlet distribution encode the results of the dice rolls. In essence, a SL opinion is formed through a series of observations that equate to tabulating the results of a number of independent rolls of the dice.

In [5], SL was expanded to handle the case that the result of the die roll is not observed. Rather a partial observation is provided that is the likelihood of a given die roll given the features that are measured from some type of physical sensor or a human generated report. The boldness¹ that can be expressed in the likelihood is a function of the quality of information inherent in the measured features [7].

This paper expands SL further to accommodate the veracity of the likelihood calculations themselves. For instance, the likelihood calculations could be the result of learned distributions from a labeled training set, and the certainty in reported likelihood values is a function of the number of training samples and the similarity between the environmental collections conditions for the training and testing data. Alternatively, one may only be able to use heuristics to approximate the likelihood. Finally, the reported likelihoods could come from another source that may be intentionally obfuscating the reported likelihoods to allow for the consumer to make certain inferences but not allow the consumer to infer other private information [8]. This work is the first step to rigorously understand how SL should reason over uncertainty in the reported likelihoods for a given partial observation.

This paper is organized as follows. Section II reviews belief updates in SL for fully visible observations, and Section III reviews the expansion of SL to incorporate partial measurement. Then, Section IV expands SL further to incorporate uncertainty in the likelihood values. Simulations demonstrate the potential utility of incorporating knowledge about the likelihood uncertainty in Section V. Finally, Section VI provides concluding remarks.

II. SUBJECTIVE LOGIC

SL is a probabilistic logic to represent one's belief in a set of K mutually exclusive assertions and the uncertainty in these beliefs [1]. Formally, SL considers a frame of K mutually exclusive singletons by providing a belief mass b_k for each singleton $k = 1, \dots, K$ and providing an overall uncertainty mass of u . These $K + 1$ mass values are all non-negative and sum up to one, i.e.,

$$\begin{aligned} u + \sum_{i=1}^K b_i &= 1, \\ \text{s.t. } u \geq 0 \text{ and } b_i \geq 0 \text{ for } i &= 1, \dots, K. \end{aligned} \quad (1)$$

SL also includes a base rate probability a_k for each singleton and a non-informative prior weight W . The collection of all the parameters forms the multinomial opinion. The base rate values represent initial (or prior) information about the probability of a singleton emerging for any given observation.

¹By boldness, we mean the degree for which the likelihood of one class (or singleton) is larger than that of the other classes.

The inclusion of the belief and uncertainty values along with the base rates and non-informative prior weight represent the accrued evidence regarding the probability of any singleton appearing in an observation. Specifically, these values map to a Dirichlet distribution for the possible probability mass function (pmf) that is controlling how singletons appear in observations. The parameters of the Dirichlet distribution α^2 are related to the multinomial opinion values via

$$\alpha_k = \frac{Wb_k}{u} + Wa_k. \quad (2)$$

Likewise, using (1), solving for b_k and u in (2) for $k = 1, \dots, K$, leads to the mapping of α to the multinomial opinions

$$u = \frac{W}{\sum_{i=1}^K \alpha_i}, \quad (3a)$$

$$b_k = \frac{u}{W}(\alpha_k - Wa_k). \quad (3b)$$

Note that a binary logic is a special case known as binary opinions, where the size of the frame is $K = 2$.

The Dirichlet distribution represents the probability distribution of the singleton likelihood probabilities p_k . The Dirichlet distribution with parameters α for the probability mass function (pmf) \mathbf{p} is

$$f_{\beta}(\mathbf{p}|\alpha) = \begin{cases} \frac{1}{B(\alpha)} \prod_{i=1}^K p_i^{\alpha_i-1} & \text{for } \mathbf{p} \in \mathcal{S}, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where

$$B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}$$

is the multinomial Beta function and the unit simplex $\mathcal{S} = \{\mathbf{p} | \sum_{i=1}^K p_i = 1\}$ is the set of admissible values of \mathbf{p} . For $K = 2$, the Dirichlet distribution is equivalent to the beta distribution. The values for the α_i 's relative to each other are equivalent to the expected value of \mathbf{p} for the Dirichlet distribution, i.e.,

$$\hat{p}_k = \frac{\alpha_k}{\sum_{i=1}^K \alpha_i}. \quad (5)$$

When the Dirichlet distribution represents the posterior, $\hat{\mathbf{p}}$ represents the minimum mean square error (MMSE) estimate of the ground truth appearance probabilities given the observations that form the beliefs. Thus, (2), (3), and (5) lead to the mapping of beliefs, uncertainty, and baseline rates to the MMSE estimates for the appearance probabilities as given by

$$\hat{p}_k = b_k + ua_k. \quad (6)$$

The Dirichlet distribution peaks near its mean value (5).³ The scaling of the Dirichlet parameters,

$$s = \sum_{i=1}^K \alpha_i, \quad (7)$$

²A boldface term α is a vector whose k -th element is α_k .

³As the Dirichlet precision increases to infinity, the peak and mean become arbitrarily close to each other.

represents the “spread” or variance of the Dirichlet distribution around its peak. Equivalently, it represents the *strength* in the confidence of the mean (or the MMSE estimate) to characterize the actual ground truth for \mathbf{p} . This value is commonly referred to as the precision parameter. As s increases, the peak becomes higher and narrower. In the limit, as $s \rightarrow \infty$, the Dirichlet converges to a Dirac delta function. Clearly by (3a), the precision value is inversely proportional to uncertainty.

The fusion of two subjective opinions consists of mapping opinions into the Dirichlet parameters, summing up the parameters while taking into account not to double-count the baseline rates, and then mapping back into the multinomial opinion space [2]. This method for fusing implies that subjective opinions are formed by observations that increment the Dirichlet parameters so that fused opinions account for all these observation increments. When the observation is the singleton that appears, the updates in SL are clear. Since the singleton appearance is drawn from the multinomial distribution, and the current belief is represented by a Dirichlet distribution, which is the conjugate prior of the multinomial, the posterior is also Dirichlet. When the k -th singleton is observed to appear, then the parameters for the updated Dirichlet distribution is known to be

$$\boldsymbol{\alpha}^+ = \boldsymbol{\alpha} + \mathbf{e}_k, \quad (8)$$

where \mathbf{e}_k is the indicator vector whose k -th element is one and whose other elements are all zero. Then $\boldsymbol{\alpha}^+$ can be inserted into (3) to obtain the updated beliefs \mathbf{b}^+ and \mathbf{u}^+ . Overall, a multinomial opinion is formed by simply counting the occurrences of singletons to maintain the Dirichlet parameters, and equivalently, the multinomial opinion values. Typically, the prior weight $W = 2$. It represents the strength of the prior in influencing updated beliefs relative to the observation.

The many operations that exist in SL for multinomial or just for binary opinions are not completely amenable to a mapping to the Dirichlet distribution in the sense of fusion and updates from observations. One example is the “and” or multiplication operation for binary opinions [9]. SL is a tractable framework, but it approximates belief propagation via parameters of a Dirichlet distribution. For any operation in SL, the operands are assumed to follow the Dirichlet distribution. A Dirichlet distribution is fitted to the output of the operation in a manner that preserves the mean. However, to maintain the properties of SL, the variance is approximated. In essence, the values of α_k ’s relative to each other are maintained. However, the sum of the α_k ’s is approximated. By (3a), the Dirichlet precision is inversely proportional to the uncertainty. These principles for handling mathematical operation in SL are used in the next two sections to add the measurement likelihood update operation into the SL framework.

III. MEASUREMENT UPDATES FOR SL

Usually, it is not possible to update beliefs in singletons by directly observing the singletons in an event. Rather, a measurement of the event is made that is statistically related to the occurrence of the singleton. The measurement forms

a feature vector $\mathbf{x} \in \mathbb{R}^d$. An observer transforms the feature vector in a likelihood vector that represents how likely the feature vector was caused by the occurrence of each of the singletons (or classes) in the frame. Either the observer does this by experience or it employs assistance from computational classifiers that learn how to represent the likelihood from a set of training data. In either case, one assumes that in this section, the observer or classifier is able to determine the correct likelihoods for each class, which is simply the value of the conditional distributions corresponding to the feature vector \mathbf{x} , i.e.,

$$l_i = f(\mathbf{x}|z = i), \quad (9)$$

where $f(\cdot|z = i)$ is the probability density function (pdf) for the measured features conditioned on the appearance of the i -th class where $1 \leq i \leq K$. In [5], the SL belief update method for partial observations that are reported as class likelihoods was developed. This section reviews that development.

A. Naïve Belief Update

The naïve approach for the partial observation update is to spread the mass of the Dirichlet update in (8) via the normalized likelihood

$$\tilde{l}_k = \frac{l_k}{\sum_{i=1}^K l_i}. \quad (10)$$

so that

$$\alpha_k^+ = \alpha_k + \tilde{l}_k. \quad (11)$$

For the case of a visible update where the value of z is known, i.e., $\mathbf{l} = \mathbf{e}_i$, then (11) simplifies to (8). While this naïve approach can be viewed as a generalization of the visible observation update, it does not yield a posterior Dirichlet distribution that is a good fit to the actual posterior distribution of the observation probabilities \mathbf{p} .

B. Likelihood Update to Approximate the Posterior

The likelihood update determines the posterior observation probabilities \mathbf{p} given the current subjective opinion and measurement. Then one fits a Dirichlet density to the posterior in order to approximate the updated subjective opinion. This derivation starts with the joint pdf of the feature, i.e., the partial observation, and the observation probabilities conditioned on the current multinomial opinion, which is

$$\begin{aligned} f(\mathbf{x}, z = i, \mathbf{p}|\boldsymbol{\alpha}) &= f(\mathbf{x}|z = i)\text{prob}(z = i|\mathbf{p})f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}), \\ &= l_i p_i f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}). \end{aligned} \quad (12)$$

Then marginalization to remove the hidden variable z leads to

$$f(\mathbf{x}, \mathbf{p}|\boldsymbol{\alpha}) = \left(\sum_{i=1}^K l_i p_i \right) f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}), \quad (13)$$

so that the posterior for the observation probabilities after the measurement update is

$$f(\mathbf{p}|\boldsymbol{\alpha}, \mathbf{x}) = \left(\sum_{i=1}^K \alpha_i \right) \frac{\left(\sum_{i=1}^K l_i p_i \right) f_{\beta}(\mathbf{p}|\boldsymbol{\alpha})}{\left(\sum_{i=1}^K l_i \alpha_i \right)}. \quad (14)$$

Note that (14) is invariant to the scaling of the likelihood. When $\mathbf{l} = \mathbf{e}_k$, i.e., the likelihood is zero for all classes except for the k -th class, then (14) simplifies to the Dirichlet distribution $f_\beta(\mathbf{p}|\alpha + \mathbf{e}_k)$, which means that the observation of the target class is fully observable. On the other hand, when all classes have equal likelihoods, i.e., $\mathbf{l} = \mathbf{1}$,⁴ (14) simplifies to $f_\beta(\mathbf{p}|\alpha)$, which means the updated beliefs are equivalent to the previous belief. In other words, the measurement is vacuous for the case of equal likelihoods. Clearly, the naïve approach given by (11) is not properly updating beliefs for the vacuous case.

The next step is to approximate the posterior by the Dirichlet distribution. The following theorem helps to determine the best Dirichlet approximation of the posterior.

Theorem 1. *For a Dirichlet distribution and the posterior distribution given in (14) to exhibit identical first order moments and to exhibit an identical second order moment for the marginal of the k -th element, then the parameters of that Dirichlet distribution are given by*

$$\tilde{\alpha}_k = s_k^+ \frac{\alpha_k + \frac{\alpha_k l_k}{\sum_{j=1}^K l_j \alpha_j}}{1 + \sum_{j=1}^K \alpha_j} \quad (15)$$

for $k = 1, \dots, K$, where the precision is

$$s_k^+ = \frac{1 + \sum_{j=1}^K \alpha_j}{1 + \frac{(2 + \sum_{j=1}^K \alpha_j)}{(\sum_{j=1}^K \alpha_j l_j)((1 + \tilde{\alpha}_k)l_k^{-1} + (1 + \alpha_k)\bar{l}_k^{-1})}}, \quad (16)$$

$$\bar{\alpha}_k = \sum_{i \neq k} \alpha_i, \quad \text{and} \quad \bar{l}_k = \frac{1}{\bar{\alpha}_k} \sum_{i \neq k} \alpha_i l_i. \quad (17)$$

Proof: See Appendices A and B in [5]. ■

Note that $\bar{\alpha}_k$ and \bar{l}_k represent the total Dirichlet precision and average likelihood, respectively, associated to the complement of the k -th singleton in the frame. When $K = 2$, it is easy to verify that $s_1^+ = s_2^+$, because $\bar{l}_1 = l_2$, $\bar{\alpha}_1 = \alpha_2$, $\bar{l}_2 = l_1$, and $\bar{\alpha}_2 = \alpha_1$. In general, $s_k^+ \neq s_j^+$ for $k \neq j$, and it not possible to select the precision so that all the second order moments of the Dirichlet approximation match those of the posterior. In any event, the larger the updated precision, the larger the updated Dirichlet parameters (see (7)).

As shown in [5], selection of even the largest value of s_k^+ as the updated precision will usually lead to the fact that one or more of the updated parameters actually decrease in value. At best, the smallest change in the updated parameters is zero. Before the update of too many observations, it is possible that the decrease in one of the parameter values leads to a negative subjective belief (see (3b)). In [5], this issue was avoided by selecting the precision to be large enough to avoid any decrease in the updated Dirichlet parameters. However, we have since discovered that a larger than necessary precision unnecessarily thwarts the influence of new observations. Better performance is obtained by only increasing the magnitude of

the precision to avoid negative subjective beliefs. Thus, the updated Dirichlet parameters are

$$\alpha_k^+ = s^+ m_k, \quad (18a)$$

$$m_k = \frac{\alpha_k + \frac{\alpha_k l_k}{\sum_{j=1}^K l_j \alpha_j}}{1 + \sum_{j=1}^K \alpha_j}, \quad (18b)$$

$$s^+ = \max \left\{ \frac{1}{K} \sum_{j=1}^K s_j^+, \frac{W a_k}{m_k} \right\}, \quad (18c)$$

where s_k^+ is given by (16). Algorithm 1 summarizes the likelihood update process for SL multinomial opinions.

Algorithm 1 Likelihood update for SL

Input: SL Multinomial opinion and likelihood \mathbf{l}

Output: Updated SL multinomial opinion

- 1) Transform multinomial opinion into Dirichlet parameters via (2).
 - 2) Update Dirichlet parameters via (16) and (18).
 - 3) Convert updated Dirichlet parameters into updated multinomial belief via (3).
-

It is shown in [5] that the increase in precision is bounded, i.e., $0 \leq s^+ - s \leq 1$. The lower bound occurs for the vacuous case where $\mathbf{l} = \mathbf{1}$. In essence, likelihood provides no information, and by (18), the subjective belief remains unchanged, i.e., $\alpha^+ = \alpha$. The upper bound occurs for the fully observable case when $\mathbf{l} = \mathbf{e}_k$ where the likelihood means the the feature values indicate without a doubt that the k -th class was observed. Then, the k -th Dirichlet parameter increments by a one while other parameters remain the same. These two cases represent the largest and smallest entropy values for the likelihood. As demonstrated in [5], as the entropy of the likelihood decreases, i.e., the likelihood is bolder in espousing a given class, the corresponding observation is increasing the precision of the updated Dirichlet parameters and lowering the uncertainty of the updated subjective belief.

IV. UNCERTAIN LIKELIHOOD UPDATES

This sections considers how to modify the belief update in (18) when the likelihood is reported with some degree of uncertainty. In practice, the likelihood calculated by an observer is not completely accurate as discussed in the introduction. For the remainder of the paper, it is convenient to collapse the likelihood onto the unit simplex \mathcal{S} . In (18), the likelihood \mathbf{l} encodes the observation \mathbf{x} as a K -dimensional vector. In fact, the magnitude of the likelihood vector provides no information as (18) is invariant to scalings of \mathbf{l} . Thus, a partial observation can be summarized as the normalized likelihood vector $\tilde{\mathbf{l}}$ (see (10)). As long as the normalized likelihood is correctly calculated, one's belief can be properly updated. For the remainder of the paper, we will use the term likelihood to actually refer to the normalized likelihood for the sake of brevity. The assumption made in this section is

⁴1 is the vector whose elements are all one.

that the distribution of the reported likelihood \tilde{l}_r conditioned on the latent actual likelihood \tilde{l} , i.e., $f(\tilde{l}_r|\tilde{l})$, is known, and uncertainty is expressed as the “spread” of this distribution. Furthermore, it is assumed that the reported likelihood conditioned on the actual likelihood is independent of the class appearance z and the appearance probabilities \mathbf{p} , i.e.,

$$f(\tilde{l}_r|\tilde{l}, z, \mathbf{p}) = f(\tilde{l}_r|\tilde{l}). \quad (19)$$

Given this distribution, this section outlines how to compute the posterior distribution of the appearance probabilities given the reported likelihood.

To compute this posterior distribution, the following theorem is useful.

Theorem 2. *The distribution of the true likelihood conditioned on the latent class appearance satisfies the following identity*

$$f(\tilde{l}|z = i) = g(\tilde{l})\tilde{l}_i,$$

where

$$g(\tilde{l}) = \sum_{i=1}^K f(\tilde{l}|z = i).$$

Proof: Consider the set of observations whose true likelihood is \tilde{l} , i.e.,

$$\mathcal{X}(\tilde{l}) = \left\{ \mathbf{x} : \frac{f(\mathbf{x}|z=i)}{\sum_{k=1}^K f(\mathbf{x}|z=k)} = \tilde{l}_i, \text{ for } i = 1, \dots, K \right\}.$$

Now,

$$\begin{aligned} f(\tilde{l}|z = i) &= \int_{\mathcal{X}(\tilde{l})} f(\mathbf{x}|z = i) d\mathbf{x}, \\ &= \int_{\mathcal{X}(\tilde{l})} \left(\sum_{k=1}^K f(\mathbf{x}|z = k) \right) \tilde{l}_i d\mathbf{x}, \\ &= \tilde{l}_i \sum_{k=1}^K \int_{\mathcal{X}(\tilde{l})} f(\mathbf{x}|z = k) d\mathbf{x}, \\ &= \tilde{l}_i g(\tilde{l}). \end{aligned}$$

The theorem states that the distribution of the likelihood given a specific class appearance is simply a scaled version of the likelihood of that class. If the observable is the likelihood and not the observation \mathbf{x} , then $f(\tilde{l}|z = i)$ replaces $f(\mathbf{x}|z = i)$ in (12). In essence, the theorem states that the posterior distribution of \mathbf{p} conditioned on the likelihood is exactly the same as the posterior distribution conditioned on \mathbf{x} since $f(\tilde{l}|z)$ is simply scaled version of the likelihood. In other words, the likelihood contains the same discriminative information as \mathbf{x} .

The scalar multiplier $g(\tilde{l})$ is inconsequential when the reported likelihood is known to be the true likelihood. However, for the more practical and general case that the observed reported likelihood is only statistically dependent on the true likelihood, then this multiplier term does affect the posterior distribution as will be seen. This multiplier encodes the phenomenology of how the likelihood is generated. If the belief update is based upon the output of a classifier, the observer

updating his/her belief must understand the distributions of the features used by the classifier, i.e., \mathbf{x} conditioned on the appearance states z . In most cases, the observer will not have a more accurate understanding of these distributions as the classifier. Therefore, in most cases $g(\tilde{l})$ is unknown to the observer. It is reasonable to consider the non-informative multiplier $g(\tilde{l}) = 1$ when the observer does not have access to the phenomenology of how the features \mathbf{x} relate to z . In the remainder of the paper, the use of the correct value for $g(\tilde{l})$ means that the feature/class likelihood model is employed. When the likelihood model is not utilized, $g(\tilde{l})$ is set to one. As will be seen in the simulations section, the shape of $g(\tilde{l})$ changes greatly if the classes are well separated or not in the feature space.

The observable is actually the reported likelihood, which statistically depends only on the true likelihood (see (19)). Likewise, the true likelihood depends only on the class appearance z , and the class appearance depends on the appearance probability. In short, the joint probability of all quantities is

$$f(\tilde{l}_r, \tilde{l}, z, \mathbf{p}|\boldsymbol{\alpha}) = f(\tilde{l}_r|\tilde{l})g(\tilde{l})\tilde{l}_z p_z f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}). \quad (20)$$

Through marginalization of the latent terms \mathbf{l} and z , the posterior distribution of \mathbf{p} given the reported likelihood is proportional to

$$f(\tilde{l}_r, \mathbf{p}|\boldsymbol{\alpha}) = \left(\sum_{z=1}^K \tilde{l}_z p_z \right) f_{\beta}(\mathbf{p}|\boldsymbol{\alpha}), \quad (21)$$

where

$$\tilde{l}_z = f(\tilde{l}_r|z) = \int_{\mathcal{S}} f(\tilde{l}_r|\tilde{l})g(\tilde{l})\tilde{l}_z d\tilde{l} \quad (22)$$

is the class likelihood of the reported likelihood. Comparing (13) with (21), it clear that \tilde{l} replaces the role of \mathbf{l} . In other words, the computation of the belief updates for a reported likelihood constitutes the transformation of the reported likelihood via (22) into an effective likelihood for insertion into the update equations.

To illustrate the general report model, we consider models for $f(\tilde{l}_r|\tilde{l})$ that simply represent that the reported likelihood is an estimate of the actual likelihood within a given level of uncertainty u_r . The uncertainty value u_r is a number between zero and one. Specifically, the reported likelihood is treated as a sample from the Dirichlet distribution with parameters $\frac{W}{u_r}\tilde{l}$. Thus, the conditional mean is the true likelihood. As $u_r \rightarrow 0$, the conditional distribution approaches a Dirac delta function centered on the true likelihood. As u_r increases, the “spread” of the distribution increases. The transformation of the reported likelihood to the effective likelihood requires a $K-1$ dimensional integration in (22). At this time, we do not know if the employment of a Dirichlet form for the conditional distribution affords a closed form formula for the transformation that circumvents the need for numerical integration.

In this paper, the transformation in (22) is accomplished via numerical integration. Thus, the paper investigates the binary

$K = 2$ case. Then the normalized effective likelihood is given by

$$\tilde{\mathbf{l}} = \begin{bmatrix} \frac{h_1(\tilde{\mathbf{l}}_r)}{h_a(\tilde{\mathbf{l}}_r)} \\ 1 - \frac{h_1(\tilde{\mathbf{l}}_r)}{h_a(\tilde{\mathbf{l}}_r)} \end{bmatrix}, \quad (23)$$

where

$$h_1(\tilde{\mathbf{l}}_r) = \int_0^1 \frac{(\tilde{l}_{r,1})^{\frac{W}{u_r}y-1} (\tilde{l}_{r,2})^{\frac{W}{u_r}(1-y)-1}}{B([\frac{W}{u_r}y, \frac{W}{u_r}(1-y)]^T)} g^*(y) dy, \quad (24a)$$

$$h_a(\tilde{\mathbf{l}}_r) = \int_0^1 \frac{(\tilde{l}_{r,1})^{\frac{W}{u_r}y-1} (\tilde{l}_{r,2})^{\frac{W}{u_r}(1-y)-1}}{B([\frac{W}{u_r}y, \frac{W}{u_r}(1-y)]^T)} g^*(y) dy, \quad \text{and} \quad (24b)$$

$$g^*(y) = g([y, 1-y]^T). \quad (24c)$$

Algorithm 2 summarizes the process to update an SL multinomial opinion using an uncertain reported likelihood.

Algorithm 2 Uncertain likelihood update for SL

Input: SL multinomial opinion, reported likelihood \mathbf{l}_r , and feature/class likelihood model $g(\tilde{\mathbf{l}})$

Output: Updated SL multinomial opinion

- 1) Normalize reported likelihood via (10).
 - 2) Transform normalized reported likelihood into effective likelihood via (22) using numerical integration.
 - 3) Update SL opinion via Algorithm 1 using the effective likelihood.
-

V. SIMULATIONS

This section uses simulations to compare various methods to update SL beliefs to demonstrate the utility of compensating for the uncertainty associated to the reported likelihoods. The simulations consider a simple exemplar case where the features are $\mathbf{x} \in \mathbb{R}^K$. For each observation, the appearance of a given class is drawn from a multinomial distribution \mathbf{p} . Given the appearance of the k -th singleton, the feature vector represents a sample from a Gaussian distribution with mean \mathbf{e}_k and covariance $\sigma^2 I$, i.e.,

$$f(\mathbf{x}|z = k) = \frac{1}{(\sqrt{2\pi\sigma^2})^K} \exp \left\{ -\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{e}_k\|^2 \right\}. \quad (25)$$

Once the feature vector is generated, the class likelihood is computed via (9) and normalized to generate the ground-truth values $\tilde{\mathbf{l}}$. Then, the reported likelihood is drawn from a Dirichlet distribution with parameters $\frac{W}{u_r} \tilde{\mathbf{l}}$ to simulate the output of the classifier. Note that in the feature space, the distances between the clusters representing the K classes, e.g., Mahalanobis distance between class centroids, are all equal. In the simulations, the number of classes is $K = 2$.

The σ^2 parameter controls the separability of the classes in the feature space. When σ^2 is small, the classes are well separated. This means that the true likelihoods are usually bold, i.e., only one element of $\tilde{\mathbf{l}}$ is large. As σ^2 grows, the class separability decreases. Given the model used in these simulations, the $g(\tilde{\mathbf{l}})$ term can be calculated analytically (see

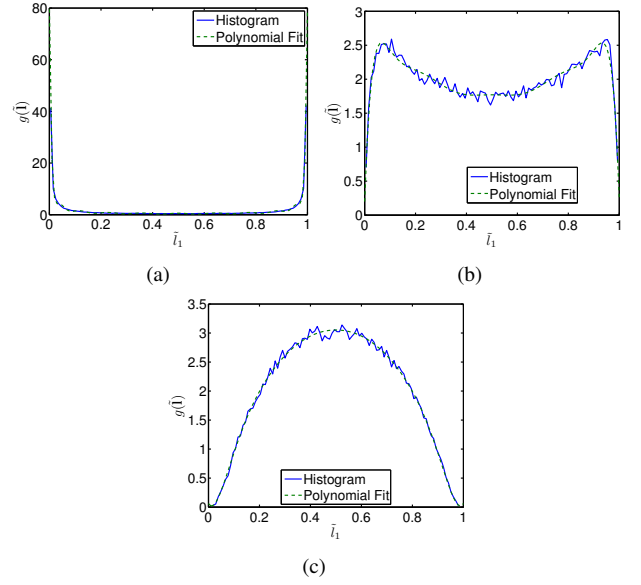


Figure 1. The estimated $g(\tilde{\mathbf{l}})$ distributions and polynomial fits to those estimates when the class spreads are (a) $\sigma^2 = 0.25$, (b) $\sigma^2 = 1$, and (c) $\sigma^2 = 2.25$.

proof of Theorem 2), but this calculations is difficult. Instead, the term is computed numerically via Monte Carlo simulations where 10,000 simulated true likelihoods are generated when the class appearance probabilities are all equal. The histogram of likelihood values is used to estimate $g(\tilde{\mathbf{l}})$. Figure 1 shows the estimated $g(\tilde{\mathbf{l}})$ for $\sigma^2 = 0.25, 1$, and 2.25 , respectively. The figure also plots polynomial fits to the estimated $g(\tilde{\mathbf{l}})$. The polynomial fits represent the feature/class likelihood models that are used in the numerical integration of (24). Since $K = 2$, the likelihoods are uniquely represented by the first element \tilde{l}_1 since $\tilde{\mathbf{l}} = [\tilde{l}_1, 1 - \tilde{l}_1]^T$. It is clear in the figure that as the the class separability decreases, the mass in the histograms migrate from the edges, i.e., bold likelihoods, to the center, i.e., vacuous likelihoods. Clearly, the shape of $g(\tilde{\mathbf{l}})$ is sensitive to the class separability in the feature space. The polynomial approximations to the histograms are used in the likelihood transformation calculations in the remainder of this section when the feature/class likelihood models are employed.

The class separability affects the transformation of the reported likelihood to the effective likelihood in (22) (and (24) for $K = 2$). Figure 2(a) plots this transformation of the reported $\tilde{\mathbf{l}}_r = [.9, .1]^T$ over the full ranges of likelihood uncertainties u_r for the three class separability models. The figure also plots the transformation when the model is ignored. For all cases, as uncertainty increases, the entropy of the effective likelihood vector increases. As demonstrated in [5], an increase in entropy of the likelihood vector leads to lower increase in the Dirichlet strength (or SL uncertainty) in the updated belief. Figure 2(b) shows the updated SL uncertainty u when the prior SL belief is $b_1 = b_2 = 0.45, u = 0.1$ with an uniform baseline of $a_1 = a_2 = 0.5$. Clearly, as the entropy of the effective likelihood increases (or its largest

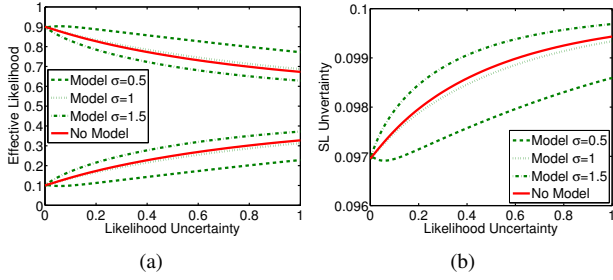


Figure 2. The effectiveness of a likelihood report of $\tilde{\mathbf{l}}_r = [0.9, 0.1]^T$ when employing and not using the feature/class model for various values of likelihood uncertainty u_r : (a) Effective likelihood values, and (b) updated SL uncertainty when prior SL opinion is $[b_1 = b_2 = 0.45, u = 0.1, \alpha_1 = \alpha_2 = 0.5]$.

element decreases), the reduction in entropy from an initial value of 0.1 decreases. When the class separability is large, the increase in entropy of the effective likelihood as likelihood uncertainty increases is slower than when class separability is smaller. This means that for a given likelihood uncertainty, the larger class separability model creates a larger decrease in the updated SL uncertainty as shown in Figure 2(b). It is interesting to note that ignoring the feature/class separability model, i.e., assume $g(\tilde{\mathbf{l}}) = 1$, provides similar behavior in the likelihood transformation as for the moderate separability model of $\sigma^2 = 1$.

The simulations compare five different possible SL belief update methods. First, the clairvoyant method uses the actual likelihood to perform the update via Algorithm 1. Because the true likelihood is used, this method serves as the unrealizable gold standard. The other methods process the reported likelihoods. The uncertainty compensation with likelihood model method performs Algorithm 2 using the true (polynomial approximation) of $g(\tilde{\mathbf{l}})$. Likewise, the uncertainty compensation without the likelihood model performs Algorithm 2 with $g(\tilde{\mathbf{l}}) = 1$. The standard method from [5], i.e., Algorithm 1, assumes certainty by treating the reported likelihood as the true likelihood. Finally, the naïve method also assumes certainty but performs the update via (11). These five methods were evaluated over each of three feature/class models and over three levels of reported likelihood uncertainty ($u_r = 0.01, 0.1$, or 1). For each of the nine feature/reported likelihood model combinations, the SL beliefs are updated one likelihood report at a time in the order of the received reports over 300 observations. Initially, the beliefs \mathbf{b} are zero so that $\boldsymbol{\alpha} = [1, 1]^T$, and the process is repeated over 1000 independent traces. For all these simulations the class appearance probabilities are $\mathbf{p} = [2/3, 1/3]^T$.

The performance results averaged over the 1000 independent traces are provide in Figure 3. The figure shows plots of the expected appearance probability for the first class appearance \hat{p}_1 (see (6)) as a function of the number of partial observations that have been integrated. Note that since $K = 2$, the relative results for the second class appearance probability are exactly the same. As one would expect, the clairvoyant method

always converges close to the true appearance probability of $p_1 = 2/3$. The naïve method never works, and the standard method (no uncertainty compensation) only converges to the true appearance probability when the likelihood uncertainty is small, i.e., $u_r = 0.01$. In fact, when the likelihood is small, all the methods but the naïve work just about equally well. This is because the reported likelihood is always very close to the true value. Usually, the standard method does not perform as well as the uncertainty compensation methods for larger uncertainty.

Uncertainty compensation using the correct feature/class separability is almost as effective as the clairvoyant method. Effectively, its performance tracks the performance of the clairvoyant method expect when the class separability is poor, i.e., $\sigma^2 = 2.25$ and the likelihood uncertainty is high, i.e., $u_r = 1$. Even in that case, the feature/class model based method is able to converge to the correct solution, but at a slower rate that the clairvoyant method. The slower convergence is due to the deep discounting of the boldness of the reported likelihood in the likelihood transformation (see Figure 2). Uncertainty compensation does not uniformly perform as well when it ignores the feautre/class model. It is usually better than no compensation at all. For moderate class separation, i.e., $\sigma^2 = 1$, its performances with and without the model are about the same as expected from Figure 2. For good class separation, i.e., $\sigma^2 = 0.25$, the no model method overcompensates the effective likelihood relative to the model method, and its expected appearance probability converges to a value slightly larger than $p_1 = 2/3$. On the other hand, when the class separation is poor, i.e., $\sigma^2 = 2.25$, the no model method undercompensates the effective likelihood, and the expected appearance probability converges to a value slightly lower than $p_1 = 2/3$. The undercompensation when ignoring the model is better than no compensation at all as the standard methods underestimates the probability more than the no model uncertainty method.

VI. CONCLUSIONS

The paper extended SL to incorporate partial observations when the reported class likelihoods are considered uncertain. The extensions is based upon approximating the results of the Bayesian update. Specifically, the current subjective belief corresponds to an equivalent Dirichlet distribution for class appearance probabilities. The reported class likelihood is treated as the observation, and the exact form for the posterior distribution is determined for the class appearance probabilities. Finally, the updated SL opinion corresponds to the parameters of a Dirichlet that best approximates the posterior by capturing the same same mean values while approximating the variances. This work build upon earlier research that determined how true class likelihood values should update the SL opinions. It turns out that considering the uncertainty of the reported likelihoods is equivalent to transforming the reported likelihoods into an effective likelihood, which then updates the SL opinions by the true likelihood update equations from the earlier work.

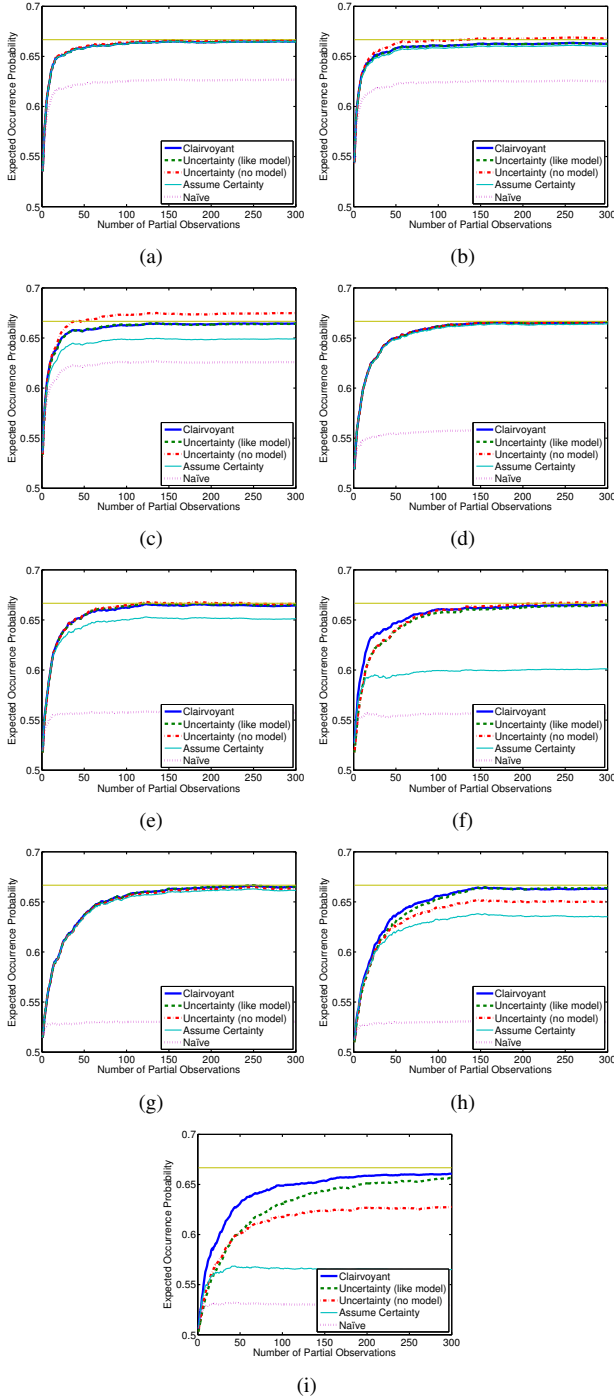


Figure 3. Expected appearance probability \hat{p}_1 when the true probability is $p_1 = 2/3$ versus number of partial observations for various SL update methods over three feature/class models and three reported likelihood uncertainty models: (a) $\sigma^2 = 0.25, u_r = 0.01$, (b) $\sigma^2 = 0.25, u_r = 0.1$, (c) $\sigma^2 = 0.25, u_r = 1$, (d) $\sigma^2 = 1, u_r = 0.01$, (e) $\sigma^2 = 1, u_r = 0.1$, (f) $\sigma^2 = 1, u_r = 1$, (g) $\sigma^2 = 2.25, u_r = 0.01$, (h) $\sigma^2 = 2.25, u_r = 0.1$, and (i) $\sigma^2 = 2.25, u_r = 1$.

Simulations demonstrated the effectiveness of the uncertainty aware update method. Actually, the uncertainty aware method requires model knowledge of how the likelihoods are generated from the actual appearances of the classes. In other words, the exact method requires knowledge of the statistical distribution of the observed features conditioned on the latent class appearance to determine the typical distribution of likelihood values. Another uncertainty aware method is proposed that ignores this feature/class model. The simulations show that ignoring the model can affect performance in terms of inferring the underlying appearance probabilities. However, even when ignoring this model, uncertainty aware compensations is usually better and never worse than taking the reported likelihood values as truth.

It would be interesting to determine if some rudimentary knowledge about the class separability afforded by the features can help in the transformation of the reported likelihood into the effective likelihood. In effect, knowledge of good class separability should simply retard the increase in the entropy of the effective likelihood as likelihood uncertainty increases. Conversely, knowledge of poor class separability should accelerate the increase in the entropy of the effective likelihood as likelihood uncertainty increases. Furthermore, the robustness of the uncertainty aware updates should be studied in terms of how well they maintain performance when the assumed model that generates the reported likelihood model does not match the actual model. Future work also needs to consider more efficient processing methods that avoids a $K - 1$ -dimensional numerical integration. Finally, we are interested in understanding how intentional obfuscation of the true likelihood can control the inferencing performance of SL.

REFERENCES

- [1] A. Jøsang, "Subjective logic," Jul. 2011, draft book in preparation.
- [2] A. Jøsang, S. Marsh, and S. Pope, "Exploring different types of trust propagation," in *Proc. of the 4th International Conference on Trust Management (iTrust)*, Pisa, Italy, May 2006.
- [3] A. Jøsang, "Conditional reasoning with subjective logic," in *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2005)*, Barcelona, Spain, Jul. 2005.
- [4] —, "The consensus operator for combining beliefs," *Artificial Intelligence Journal*, vol. 142, no. 1–2, pp. 157–170, Oct. 2002.
- [5] L. M. Kaplan, S. Chakraborty, and C. Bisdikian, "Fusion of classifiers: A subjective logic perspective," in *Proc. of the IEEE Aerospace Conference*, Big Sky, MT, Mar. 2012.
- [6] P. Shokarian, V. S. Subrahmanian, and M. L. Sapino, "GAPs: Geospatial abduction problems," *ACM Trans. on Intelligent Systems and Technology*, vol. 3, no. 1, Oct. 2011.
- [7] C. Bisdikian, L. M. Kaplan, M. B. Srivastava, D. J. Thornley, D. Verma, and R. I. Young, "Building principles for a quality of information specification for sensor information," in *Proc. of the International Conference on Information Fusion*, Seattle, WA, Jul. 2009.
- [8] S. Chakraborty, H. Choi, and M. B. Srivastava, "Demystifying privacy in sensory data: A QoI based approach," in *Proc. of the IEEE IQ2S 2011*, Mar. 2011.
- [9] A. Jøsang and D. McAnally, "Multiplication and comultiplication of beliefs," *International Journal of Approximate Reasoning*, vol. 38, no. 1, pp. 19–51, 2004.